Chapter 7: Data Collection and Capture for Housing Unit Addresses

[This page intentionally left blank]

Table of Contents

Figures

Tables

[This page intentionally left blank]

Chapter 7: Data Collection and Capture for Housing Unit Addresses

7.1. Overview

The data collection operation for housing unit (HU) addresses consists of four modes: internet, paper, telephone, and personal visit. For most HU addresses, the first phase includes a mailed request to respond via internet, followed later by an option to complete a paper questionnaire and return it by mail. If the Census Bureau does not receive a response or if the household refuses to participate, the address may be selected for computer-assisted personal interviewing (CAPI).

The American Community Survey (ACS) divides the overall sample into monthly independent samples which are called panels. Data collection for each panel lasts for three months, with paper and internet returns accepted during this entire period, as shown in Figure 7-1. This two-phase process operates in continuously overlapping cycles so that, during any given month, three samples are in the internet/paper response (self-response) phase and one panel is in the CAPI phase.

Figure 7-1: ACS Data Collection Timeline (Example for selected panels)
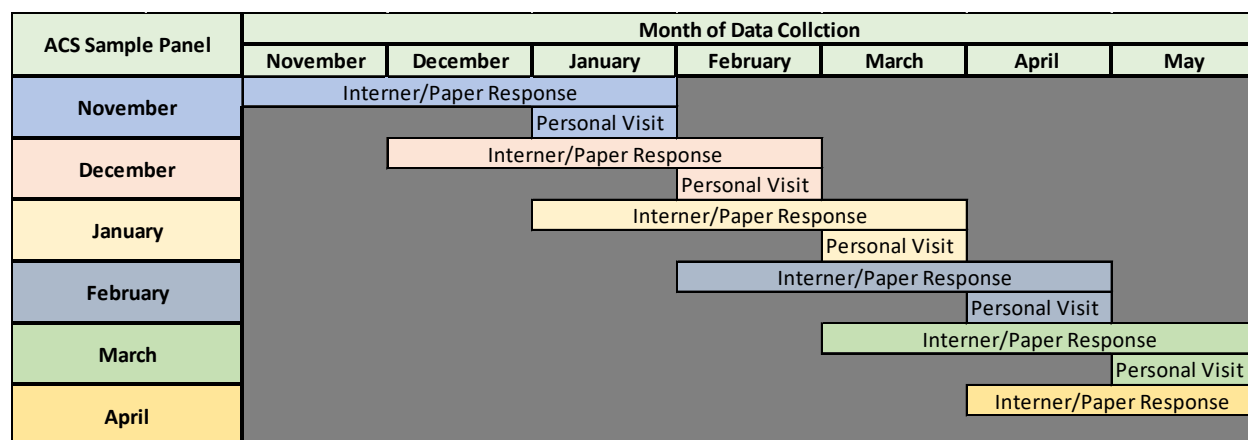
For information on ACS response rates, visit: Response Rates | American Community Survey | U.S. Census Bureau. For more information on the distribution of response mode by item, visit: https://www.census.gov/library/visualizations/interactive/acs-collection.html.

7.2. Self-Response Phase

Self-response by paper or internet is the least expensive method of data collection, and the success of the program depends on high levels of self-response. Sample addresses are reviewed to determine whether the available information is sufficient for mailing. The requirement for a "mailable" address in the United States is met if there is either a complete city-style or rural route address. A complete city-style address includes a house number, street name, town or city name, state, and ZIP Code. A complete rural-route address includes a rural-route number, box

number, town or city name, state, and ZIP Code. About 99 percent of the 2022 sample addresses in the United States met these criteria and were designated as mailable.

The requirement for a mailable city-style Puerto Rico address is the same as Stateside. Mailing addresses in Puerto Rico generally incorporate the urbanization name (a geographic area roughly equivalent to a neighborhood), which creates a four-line address. Use of the urbanization name eliminates the confusion created when street names are repeated in adjacent communities. About 70 percent of the addresses in Puerto Rico were considered mailable in 2022.

Examples of unmailable addresses include those with only physical descriptions of an HU and its location, or with post office (P.O.) box addresses, as well as addresses missing place names and zip codes. P.O. box addresses are considered unmailable because of the unknown location of the HU using the P.O. box. Addresses missing zip codes are considered unmailable when the place name is also missing. HU addresses not meeting one of the completeness criteria are still included in the sample frame and are eligible for the Internet phase, but they bypass the paper response phase.

ACS Mail Contact Strategy[1]

Because a high level of self-response is critical, the ACS sends up to five mailings to encourage recipients to respond to the survey via the internet or to return a paper questionnaire. Figure 1 outlines the ACS mail contact strategy.

Figure 7-2. ACS Mail Contact Strategy
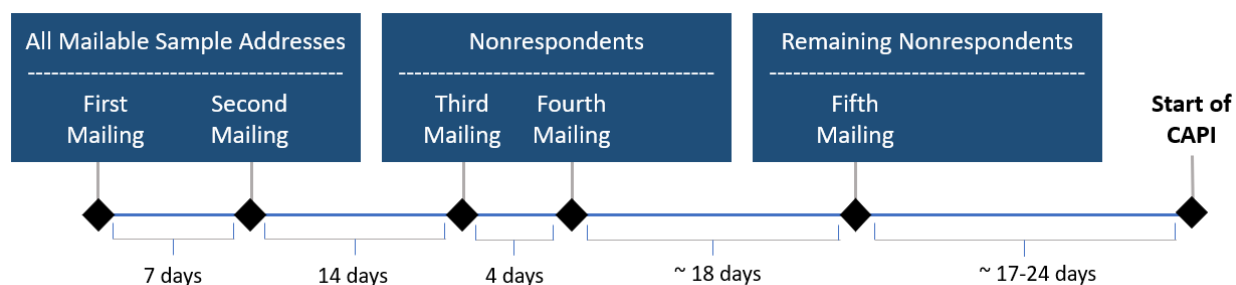
ACS materials for U.S. addresses are sent in English. Respondents can request Spanish mailing packages via telephone questionnaire assistance (TQA). All mailable sample addresses are sent the first two mailings: an initial letter and a reminder letter. A paper questionnaire package and reminder postcard are sent to sample addresses when there is no response about three weeks after the initial mailing. Households that have not responded are sent another letter with a due date the following month. Details of each mailing are provided below.

---

[1] See below for information on the Puerto Rico Community Survey Mail Strategy.

Initial Mail Letter. The first mailing is a pressure-sealed letter, signed by the director of the Census Bureau. It informs the householders they are in sample and encourages them to go online to complete the survey as soon as possible. Login information is provided in the letter. The letter explains the purpose of the ACS and how the data are used. Finally, a toll-free telephone number is included for respondents if they have questions or need help completing the questionnaire. The letter is mailed on the last Thursday of the month prior to the self-response month or on the first Thursday of the self-response month if the last day of the month is a Wednesday. The initial mailing letter is printed in-house at the Census Bureau's National Processing Center (NPC), using print-on-demand technology to merge text and addresses and user ID.

First Reminder Letter. The second mailing is a pressure sealed letter, signed by the director of the Census Bureau. The letter is mailed on Thursdays, a week after the initial mail letter, and reminds respondents to complete the survey via the internet. The reminder letter is printed in-house at the NPC, using print-on-demand technology to merge text and addresses and user ID.

Paper Questionnaire Package. Responding addresses are removed from the address file after the second mailing to create a new mailing universe of nonrespondents; these addresses are sent the third and fourth mailings.[2] The third mailing is a package containing a paper questionnaire along with a follow-up letter and a postage-paid return envelope. Signed by the director of the Census Bureau, it reminds the household of the importance of the ACS and asks them to respond soon (either online or by returning the paper questionnaire). The paper questionnaire package is mailed on Thursdays, about three weeks after the initial mail package.

The NPC assembles and mails the packages for the selected addresses. All the components of the mail packages are printed under contract by outside vendors. As the vendors print the materials, NPC quality control staff monitor the work and inspects materials to identify samples that do not meet contractual quality standards.

Reminder Postcard. The fourth mailing is a postcard, printed on white cardstock and reminds respondents to return their questionnaires or respond via internet. The postcard is mailed on Mondays, a few days after the paper questionnaire package, This postcard also is printed in-house at the NPC, using print-on-demand technology to merge text and addresses.

---

[2] Addresses deemed "undeliverable as addressed" (UAA) by the United States Postal Service (USPS) are also removed from the address files for subsequent mailings.

Additional Letter. After the fourth mailing, responding addresses are again removed from the address file to create a new mailing universe of nonrespondents. The remaining sample addresses are sent the fifth mailing (a more urgent final reminder letter with a due date in a pressure seal letter, signed by the director of the Census Bureau). It is mailed on Tuesdays, about seven weeks after the initial mail package. The final reminder letter is printed in-house at the NPC, using print-on-demand technology to merge text displaying the due date, address, and user ID.

The NPC is responsible for printing the pressure sealed letters and labeling the outgoing mail packages. Several months before each sample's mailings, Census Bureau headquarters staff provides an address file to the NPC for use in creating address labels for the first two mailings. An updated address file is provided to the NPC about three days before the mailing of the paper questionnaire package. This file excludes addresses from which a response was received by internet during the first two and a half weeks and addresses identified as "undeliverable as addressed" (UAA) by the U.S. Postal Service (USPS); these usually amount to about 24 percent of the sample addresses. An additional updated address file for the final reminder mailing is provided to the NPC two and a half weeks after paper questionnaire package mailings which excludes addresses from which a response was received via paper or internet.

Most paper and internet responses are received within five weeks after the initial mail package is sent, but the Census Bureau will continue to accept paper or internet responses for three months from the start of each monthly sample. After a specified cutoff date, late returns will not be included in the data set.

Mail Contact Strategy in Puerto Rico

An internet response option is not available for Puerto Rico (PR) households; therefore, a slightly different set of mailings are used in Puerto Rico. Figure 2 outlines the Puerto Rico Community Survey (PRCS) mail contact strategy.

Figure 7-3. PRCS Mail Contact Strategy
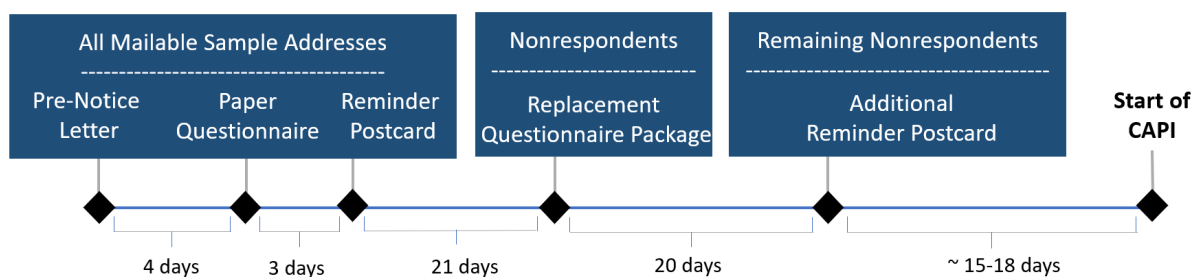
PRCS materials are sent in Spanish. Respondents can request English mailing packages, via TQA. Puerto Rico households receive a pre-notice letter prior to an initial mailing package. This letter notifies them that they've been sampled for PRCS, provides information on the importance of ACS, and instructions to respond. The initial mailing package includes a paper questionnaire rather than a pressure-sealed letter with information for responding via internet, the replacement package does not include instructions for responding via internet and there is no second reminder postcard. Unlike U.S. households, the 5th mailing for PR households is a postcard rather than a pressure-sealed letter. The timing of the replacement mail package is approximately 3 ½ weeks after the initial mail package for Puerto Rico households.

For Puerto Rico, all the components of the mail packages except the pre-notice letter and postcards are printed under contract by outside vendors. As the vendors print the materials, NPC quality control staff monitor the work and inspect materials to identify samples that do not meet contractual quality standards. The NPC is responsible for printing the letters and postcards, assembling, labeling, and mailing the outgoing mail packages. Several months before each sample's mailings, Census Bureau headquarters staff provides an address file to the NPC for use in creating address labels for the first three mailings. An updated address file is provided to the NPC about three days before the mailing of the replacement mail package. This file excludes Puerto Rico addresses from which a response was received during the first two and a half weeks and addresses identified as UAA by the USPS; these usually amount about 21 percent of the sample addresses for Puerto Rico. An additional updated address file for the final reminder mailing is provided to the NPC two and a half weeks after replacement mailings which excludes addresses from which a response was received via paper or internet.

Most paper responses are received within five weeks after the initial mail package is sent, but the Census Bureau will continue to accept paper responses for three months from the start of each monthly sample. After a specified cutoff date, late returns will not be included in the data set.

Internet Response

As mentioned earlier, the initial mailing for the ACS includes an invitation to complete the survey online.[3] The letter contains a unique code and instructions on how to get to the response web site. After logging in, the householder answers challenge questions and is provided a code that will allow them to exit and return later to complete the survey.

After completing the information for each person and the housing unit, the householder can review their data and make any needed corrections before submitting the form. Once a householder submits a form, they cannot access it anymore and the data are processed.

---

[3] PRCS does not have an internet response option.

Households at a sample address can complete an internet return at any time during the three months of data collection for the panel it is assigned. If a household has started an internet return but not completed it, data from those "partial interviews" are processed at the end of the three-month data collection period.

Paper Questionnaire Response

The 2022 ACS questionnaires are 48-page, two-color booklet-style forms. They are printed on white paper with colored ink—green for the ACS form, purple for the PRCS form. The cover of the questionnaire includes information in English and Spanish on how to obtain assistance, and information on how to respond via the internet. The questionnaire includes questions about the HU and the people living in it. Space is provided for detailed information for up to five people. Follow-up by telephone is used for households that return their questionnaires by mail and report that six or more people reside in the household (see Failed-Edit Follow-Up below for more information).

The USPS returns all completed ACS paper questionnaires to the NPC. The check-in unit receives mail deliveries two or three times each business day. Each questionnaire contains a unique bar code in the address label area. The paper returns are sent through a laser sorter, where the bar code is scanned; this allows sorting by and within monthly sample and by location. During this step, the return envelopes are opened mechanically. Staff have three days to check in a form, although usually they check in all forms they receive within one day. Each day, NPC staff transmit a file of the checked-in cases, and headquarters staff update the status of each case in the control file.

After clerks remove the forms from the return envelopes, each page of the forms is thoroughly reviewed. The clerks look for enclosed correspondence, which they forward to headquarters, if necessary. The clerks then scan the bar code on each questionnaire to officially check in the form and organize the forms into batches of 50.

Some of the forms are returned to the NPC as UAA by the USPS. UAAs occur for many reasons, including bad or unknown addresses, vacant HUs, or residents' refusals to accept mail delivery. Sample addresses that are UAAs are ineligible for the replacement mail packages. UAAs are eligible for the CAPI operations.

Data Capture of Paper Questionnaires

After the paper questionnaires have been checked in and batched into groups of 50, they are scanned and the scanned images continue through the data capture process. Checkboxes on the images are read by optical mark recognition (OMR) and numeric entries are read by optical character recognition (OCR). Any value not read with high confidence by OMR or OCR is presented to a clerk or keyer. After OMR and OCR, the images move to the data entry (keying) unit in the NPC. The keying unit has the goal of keying the responses from the questionnaires

within three weeks of receipt. Data keyers enter the information from the forms into a data capture file. Each day, NPC staff transmit a file with the keyed data, and headquarters staff update the status of each case in the control file. The NPC's data keying operation uses stringent quality assurance procedures to minimize nonsampling errors.

Data keyers move through two stages of quality assurance verification. When new keyers begin their training, they key four truth deck batches. There is no error rate calculation on the first two batches. If the keyer's field error rate for the third truth deck (the percentage of all keyed data items that are in error) is equal to or less than 1.0 percent, the keyer will be given the fourth truth deck batch to key. If the keyer's field error rate for the fourth truth deck batch is equal or less than 1.0 percent, the keyer is moved to the qualified stage. If the keyer's field error rate for the third or fourth truth deck batch is greater than 1.0 percent, the keyer is given numerous attempts to retry keying the truth deck batches. If the keyer's field error rate is greater than 1.0 percent on the third or fourth batch after numerous attempts, the keyer will not be allowed to key for the ACS.

Once the keyer has cleared training, they are qualified to key and verify batches. If these keyers exceed the production error rate of 1.0 percent, they receive timely feedback. Keyers at all levels are subject to removal from the project and administrative action if they fail to maintain a production error rate of less than 1.0 percent, but most have a much lower rate.

Telephone Questionnaire Assistance (TQA)

Respondents that call the toll-free TQA number reach an interactive voice recognition (IVR) telephone system that provides answers to questions about completing the questionnaire or assists respondents in requesting a questionnaire in another language. The TQA telephone number is listed on the questionnaire, as well as on all the letters, brochures, and postcards. Alternate TQA numbers are listed on the questionnaire for Spanish speakers and for a telecommunications device for the deaf (TDD).

When respondents call TQA, they enter the IVR system, which provides some basic information on the ACS and directions on using the IVR. Respondents may obtain recorded answers to FAQs, or they can speak directly to an agent during business hours. Respondents can furnish their ACS identification number from any of the mailing pieces, which allows them to hear a customized message about the current status of their paper or internet questionnaire. The IVR can indicate whether the NPC has received a completed survey for the sample address and, if not, can state that an ACS interviewer may call or visit. If a respondent chooses to speak directly to an agent, the agent answers the caller's questions and gives the respondent the option to complete the questionnaire over the telephone. Agents use an automated survey instrument to capture the respondent's answers. Respondents may also contact TQA staff to reset their internet questionnaire if the personal identification number (PIN) number provided when the respondent

first accessed their questionnaire online is lost and if they are unable to provide answers to their security questions.

Household members from approximately eight percent of the mailable addresses called the toll-free number and talked to a TQA agent for assistance in 2019. While the TQA agents can complete the survey over the phone, less than one percent of the mailable addresses are completed that way. All calls are logged, and the system can record up to five reasons for each call. Even though TQA interviews are conducted by telephone, they are considered paper responses because the call was initiated by the sample household upon receiving the initial mailings.

Failed-Edit Follow-Up

After the data are keyed from paper or the data are returned from the internet, the data files are processed through a computerized edit to check coverage consistency. This edit identifies cases requiring additional information. Cases that fail are eligible for the telephone failed-edit follow-up (FEFU) operation and become part of the FEFU workload if a telephone number for the sample address is available. This operation is designed to improve the final quality of completed surveys.

Cases failing the edit for coverage consistency can take three forms. First, since the ACS paper questionnaire is designed to accommodate detailed answers for households with five or fewer people, a case will fail when a respondent indicates that there are more than five people living in the household, or if the reported number of people is greater than the number of people for whom responses are provided. Second, responses that indicate the sample address is vacant or a business are also treated as coverage failures. Third, the number of data-defined fields is less than expected and the case data will not be retained for downstream processes. A new set of FEFU cases is generated each business day, and telephone center staff call respondents to obtain the missing data. The interview period for each FEFU case is three weeks.

7.3. Personal Visit Phase

The last phase of ACS data collection is the personal visit phase, or CAPI. This phase usually begins on the first day of the third month of data collection for each panel, and typically lasts for the entire month.

Two to three weeks after the fifth mailing, responding addresses are removed and unmailable and UAA addresses are added back in to create the universe of addresses eligible for CAPI. Of this universe, a subsample is selected from two categories of cases to be included in the CAPI operation. Mailable addresses with neither a paper or internet response are sampled at a rate of 1 in 2, 2 in 5, or 1 in 3 based on the expected rate of completed interviews at the tract level. Unmailable addresses are sampled at a rate of 2 in 3. All eligible addresses in Hawaiian

Homelands, Alaska Native Village Statistical areas, and a subset of American Indian areas are sent to CAPI without subsampling (U.S Census Bureau, 2020).

Mailable households in the CAPI subsample are sent a pressure sealed letter encouraging them to complete the internet questionnaire to avoid an in-person interview. The letter is mailed at the beginning of CAPI. The reminder letter is printed in-house at the NPC, using print-on-demand technology to merge text and addresses and user ID.

The CAPI operation is conducted by Census Bureau field representatives (FRs) operating from the Census Bureau's six regional offices (ROs). The sampled cases are distributed among the six ROs based on their geographic boundaries. The New York RO is responsible for CAPI data collection in Puerto Rico.

After the databases containing the sample addresses are distributed to the appropriate RO, the addresses are assigned to FRs. The FRs can conduct interviews by telephone or personal visit, using laptop PCs loaded with the survey instrument. The CAPI instrument is available in English and Spanish in the United States and Puerto Rico.

If a telephone number is available, the FR will first attempt to call the sample address. Unmailable addresses are an exception because an FR must verify the location of the address in person. The FR will call and confirm that he or she has reached the sample address. If so, the FR uses the automated instrument and attempts to conduct the interview. If an FR cannot reach a resident after calling three to five times at different times of the day during the first few days of the interview period, he or she must make a personal visit.

Approximately 70 percent of CAPI cases require an FR visit. In addition to trying to obtain an interview, a visit is needed to determine whether the HU exists and to determine the occupancy status. If an HU does not exist at the sample address, that status is documented. If an FR verifies that an HU is vacant, he or she will interview a knowledgeable respondent, such as the owner, building manager, real estate agent, or a neighbor, and conduct a "vacant interview" to obtain some basic information about the HU. If the HU is currently occupied, the FR will conduct an "occupied" or "temporarily occupied" interview. An FR conducts a temporarily occupied interview when there are residents living in the HU at the time of the FR's visit, but no resident has been living there or plans to live there for more than two months.

Collecting data on American Indian Reservations (AIRs) is unique because of the sophistication of tribal government's functionality and diverse processes across reservations. For the first visit to the AIR, the RO management staff or Tribal Specialist Staff will contact the tribal government and establish a contact person for the initial visit and subsequent visits. Each tribe will have their own protocol for conducting interviews on the reservation.

The FRs are trained to remain polite but persistent when attempting to obtain responses. They also are trained on how to handle almost any situation, from responding to a household that claims to have returned its questionnaire by mail, or responded by internet, to conducting an interview with a non-English speaking respondent.

When FRs cannot obtain interviews, they must indicate the reason. Such noninterviews are taken seriously because they have an impact on both sampling and nonsampling error. Noninterviews occur when an eligible respondent cannot be located, is unavailable, or is unwilling to provide the survey information, or when the maximum number of attempts allowed across all data collection modes has been reached. Additional noninterviews occur when FRs are unable to confirm the status of a sample HU because of restricted access to an area because of a natural disaster or nonadmission to a gated community during the interview period. Some sample cases will be determined to be ineligible for the survey. These include sample addresses of structures under construction, demolished structures, and nonexistent addresses.

Approximately 2,500-3,000 FRs conduct CAPI interviews across the United States and Puerto Rico. Interviewers have almost the entire month to complete the monthly CAPI caseload, which averaged approximately 67,000- 69,000 cases each month in 2022. Each day, FRs transmit a file with the status of all personal visit cases, and headquarters staff update the statuses on the control file.

FRs participate in a 4-day classroom training session to learn and practice the appropriate interviewing procedures. Each FR is also trained thoroughly on the importance of maintaining high data quality. Since ACS estimates extend to the block group level, and most CAPI cases are subsampled, each potential CAPI interview should be handled with care. Supervisors travel with FRs during their first few work assignments to observe and reinforce the procedures learned in training.

In addition, a sample of FRs is selected each month and supervisors reinterview a sample of their cases. The primary purpose of the re-interview program is to verify that FRs are conducting interviews and doing so correctly. Analysts also review survey paradata and response data to identify FRs who may not be collecting quality data and notify FRs of any concerns.

Data Collection in Remote Alaska

Remote areas of Alaska provide special difficulties when interviewing, such as climate, travel, and seasonality of the population. To address some of these challenges, the Census Bureau has designated some of these areas to use different procedures for ACS interviewing.

For areas of Alaska that the Census Bureau defines as remote, ACS operations are different from those operations in the rest of the country. The Census Bureau does not send mail soliciting self-response to Remote Alaska sample units, which means Remote Alaska respondents do not

receive a paper questionnaire. We also do not attempt to conduct interviews with households in Remote Alaska via Census Bureau telephone center interviewers. All interviews for Remote Alaska are conducted using personal visit procedures only, and we do not subsample for CAPI in Remote Alaska as we do elsewhere. During the personal visits, FRs can provide respondents information on how to respond to the ACS via the internet.

To allow FRs in Alaska adequate time to resolve some of the transportation and logistical challenges associated with conducting interviews in Remote Alaska areas, the normal period for interviewing is extended from one month to six months. There are two 6-month interview periods every year in Remote Alaska. The first starts in January and stops at the end of June. The second starts in July and stops at the end of December. Extending the interview periods into two 6-month periods was found to be more effective in allowing FRs to gain access to remote areas, and in finding residents of Native Villages at home who might be away frequently to participate in subsistence activities.

Table 7-1 provides a list of Remote Alaska areas and their associated interview periods.

Table 7-1. Interview Periods by Area in Remote Alaska Areas

Borough name	Interview period for the remote portion of the borough: January – June	Interview period for the remote portion of the borough: July – December
Aleutians East	X	
Aleutians West		X
Bethel		X
Bristol Bay	X	
Chugach	½	½
Copper River	½	½
Denali		X
Dillingham	X	
Haines		X
Hoonah-Angoon	½	½
Kusilvak	½	½
Lake and Peninsula	X	
Nome	½	½
North Slope	X	
Northwest Arctic	½	½
Prince of Wales-Hyder	½	½
Southeast Fairbanks	½	½
Wrangell	½	½
Yakutat	X	
Yukon-Koyukuk		X

Note: X indicates that all workload falls in the interview period. ½ indicates that half of the workload falls in the interview period.

7.4. References

U.S. Census Bureau. 2020. "Accuracy of the Data (2020)." Washington, DC, 2020, American
Community Survey Accuracy of the Data (2020) (census.gov)

Chapter 7 Revision History

Author	Date	Brief Description of Change	Version #
C. Rosol	1/30/2014	Final draft (2014 edition) to be published	2.0
D. Reese & E. Poehler	11/28/2022	Updated content and formatting.	3.0